

# 強化学習型ベイジアンネットワークを用いた分人モデルの構築

河村 将範

指導教員：小林 邦和

## 1 はじめに

本研究に至った経緯として、ヒューマン-ロボットインタラクションという研究分野に着目した。近い将来、ロボットが人と同じようにインタラクションをとることで、ロボットが人の助けとなったり、話し相手となってくれることで人々が過ごしやすい社会となることが想定される。また、人が相手に対して接し方を使い分けるといった分人という概念が平野により提唱されている [1]。本研究では、この分人という概念をモデル化し、自己組織化マップ (以降 SOM と略す) [2]、ベイジアンネットワーク (以降 BN と略す) [3]、Q 学習 (強化学習) [4]、Soft-max 行動選択法 [5] を用いて、分人モデルの構築を行った。また、ベイジアンネットワークが持つ条件付き確率表 (主に、統計データを基にして一意に設定される) を Q 学習によって学習させ、ベイジアンネットワークの動的な学習を実現した。

## 2 分人とは

人は、他人との接し方を使い分けことによってインタラクションを行っている。例として、自分にとって目上の方に対しては敬語を用い、友人に対しては普段使っているような言葉で話をする。目上、友達という簡単な分類を例に取り上げたが、人は性別や国籍、自分との関係などの要因により、話し方や接し方を使い分けている。このような概念を分人と呼ぶ。分人は 3 つに大別される。1 つ目は、見ず知らずの人と当たり障りのないインタラクションを行う社会的分人である。2 つ目は、複数の人とインタラクションを行うグループの分人である。3 つ目は、特定の人に向けた特定個人の分人である。

## 3 分人モデル

### 3.1 分人モデル概要

今回構築した分人モデルは、現在インタラクションを行っている相手の特定と、その人に関する情報を学習する分人特定部、相手や環境情報に応じて行動の選択と学習を行う行動決定部に大別される。分人モデル全体のモデル図を図 1 に示す。

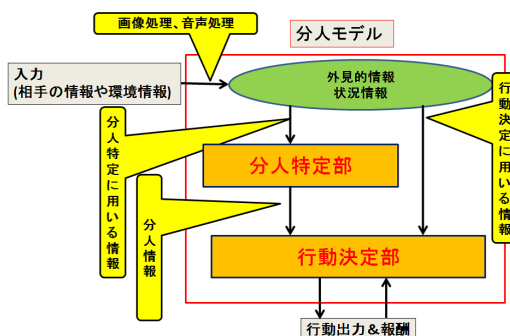


図 1 分人モデルのモデル図

図 1 の赤枠で囲まれた部分が今回提案する分人モデルである。このモデルの流れとしては、まず相手の情報や環境情報といった入力を受け取り、ロボットの画像処理、音声処理によって外見的信息や状況情報を抽出する。ここで扱う画像処理、音声処理とは、他者の特徴や環境の特徴を抽出するものである。後に行うシミュレーション環境では、相手の年齢、性別といった外見的信息や、季節、天気などの状況情報が抽出されたものとする。その後、分人特定に用いる情報と行動決定に用いる情報に区別される。分人特定に用いる情報は、主に相手の年齢や性別といった相手の外見的信息であり、行動決定に用いる情報は、主に季節、天気などの状況情報である。このように、外見的信息、状況情報は、分人特定部で用いる情報と、行動決定部で用いる情報に区別される。分人特定部によって分人が特定された後、その分人情報を行動決定部へと入力する。そのため、今回のモデルの

流れとしては、分人特定部で分人を特定した後に、その情報を用いて行動決定を行う。分人情報は、具体的には、現在インタラクションを行っている相手の分人が社会的分人であるのか、グループの分人であり、どのグループの分人なのか、または特定個人の分人であり、どの特定個人の分人なのかという情報である。この分人情報と行動決定に用いる情報を用いて、行動決定部にて行動を決定し、インタラクションの相手に出力する。その後、そのインタラクションの内容に応じて報酬を取得し、その報酬を基に分人モデルの学習を行う

### 3.2 分人特定部

分人特定部では、SOM を用いて分人の特定と学習を行う。相手の情報に関する入力と、SOM が持つユニットとの距離が最も近いユニットに振られているラベルに応じて、分人を特定する。もし、そのユニットにラベルが付けられていない場合には、社会的分人として特定したものとし、次回以降の分人の特定のために、入力情報とユニットが持つ参照ベクトルとの距離に応じてその人に対するラベルをユニットに振る。

### 3.3 行動決定部

行動決定部では、環境や行動などの要因の因果関係を BN によって表現し、行動の選択確率を Q 学習によって更新し、Soft-max 行動選択法を用いて行動選択確率に基づいた行動決定を行う。また、分人特定部の SOM 内に存在するラベルの数だけ BN は存在する。行動選択は、特定された分人に対する BN を用いて行い、その BN を更新する。BN が持つそれぞれのノードの条件付き確率は、Q 学習における Q 値を利用して算出される。

## 4 計算機シミュレーション

### 4.1 分人特定部計算機シミュレーション

本節では、分人特定部に関して、計算機シミュレーションによる性能評価を行う。

#### 4.1.1 問題設定

今回の分人特定部における計算機シミュレーションでは、4 人 (父親、母親、長男、長女) の入力情報に対して、SOM 内のユニットが学習されることをシミュレーションにより確認する。また、今回は、入力情報によって分人が学習されることを明確に視覚化するため、社会的分人を含まないものとする。

#### 4.1.2 パラメータ設定

使用する入力は、将来的に分人モデルをロボットに導入することを踏まえ、ロボットがカメラやマイクから取得したセンサー情報を前処理した入力とする。また、暗い場所、明るい場所といった要因に応じて、入力情報にノイズが入ることを想定し、1 人に対して類似した 2 パターンの入力を用意した。具体的な入力データを表 1 に示す。

表 1 入力データ

人物	(年齢, 性別, 髪型, 顔の輪郭, 体型)
父親(A)	(0.8, 0.5, 0.2, 0.1, 0.3)
	(0.8, 0.6, 0.1, 0.1, 0.7)
母親(B)	(0.5, 0.2, 0.9, 0.4, 0.3)
	(0.5, 0.3, 0.8, 0.4, 0.7)
長男(C)	(0.3, 0.7, 0.4, 0.3, 0.3)
	(0.3, 0.8, 0.5, 0.2, 0.7)
長女(D)	(0.1, 0.3, 0.7, 0.6, 0.3)
	(0.2, 0.3, 0.8, 0.6, 0.7)

また、SOM のパラメータ設定として、SOM のユニット数を  $10 \times 10$ 、割引率  $\alpha = 0.3$ 、分散  $\sigma = 1.0$  と設定した。

### 4.1.3 結果

分人特定部での計算機シミュレーションの結果を図 2 に示す。学習後のこの SOM に、母親に類似した情報として、(0.5, 0.4, 0.7, 0.4, 0.5) という情報を入力したところ、図 2 上において左上から数えて右に 6 個目、下に 6 個目のユニット (紫色のユニット) が反応し、母親であることが特定された。

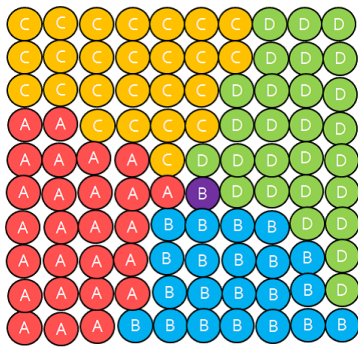


図2 分人特定部計算機シミュレーション結果

#### 4.1.4 考察

計算機シミュレーション結果より、分人がラベルによってカテゴリ分けされていることが分かる。このように、インタラクションを行った相手を学習し、SOMによって視覚的に分人の構成を捉えることができる。また、学習後のSOMに、母親に類似した情報を入力したところ、母親と特定することができた。同様に、父親、長男、長女に類似した情報を入力したところ、それぞれの人を特定することが確認できた。このように、現在インタラクションを行っている相手を特定することができる。

#### 4.2 行動決定部計算機シミュレーション

本節では、行動決定部に関して、計算機シミュレーションによる性能評価を行う。

##### 4.2.1 問題設定

今回の行動決定部における計算機シミュレーションでは、こちらから相手に”話題を提供する”というインタラクションを想定し、相手の好みに合わせた話題の提供ができるように行動を学習できることを計算機シミュレーションによって確認する。また、将来的に父親、母親、長男、長女の4人家族のホーム用ロボットに分人モデルを導入することを想定し、話題の設定を行った。話題提案における報酬の取得方法としては、相手の反応があれば大きな報酬、反応がなければ小さな報酬を得られるものとした。今回の計算機シミュレーションでは、分人特定部にて、長男とインタラクションを行っていることを特定できたものとし、話題の提供に関するBNの行動選択確率が学習されることを確認する。今回構築したBNを図3に示す。

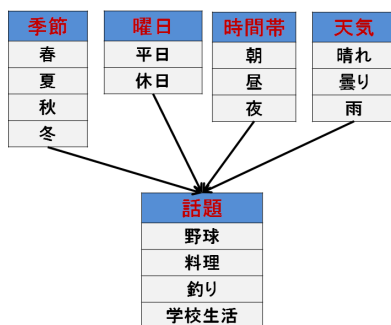


図3 話題提供に関するBN

##### 4.2.2 パラメータ設定

今回のシミュレーションでは、長男は野球と学校生活の話題には興味を示し(自己は大きな報酬を受け取ることができる)、料理と釣りの話題には興味を示さない(自己は小さな報酬を受け取ることができる)こととした。インタラクションの条件として、季節=夏、曜日=×(観測できなかった)、時間帯=朝、天気=曇りという情報を観測できたものとし、これを100回繰り返した。その後、季節=×、曜日=休日、時間帯=昼、天気=曇りという観測状態となり、この条件の下で100回のインタラクションを行った。Q値の初期値は全て0と設定し、BNのノードそれぞれが持つ条件付き確率の初期値は、確率の総和が1となる性質を利用し、ノードがもつ状態数で1を割った等確率とし、これを基準値として設定した。また、BNの条件付き確率表の更新に使用

したQ学習のパラメータを、学習率 $\alpha = 0.1$ 、割引率 $\gamma = 0.8$ 、大きな報酬を50、小さな報酬を0として設定した。

##### 4.2.3 結果

今回は、曜日が持つ2つの状態のうち、休日における話題ノードの条件付き確率表を表2に示す。

表2 休日におけるそれぞれの条件の下での話題ノードに関する条件付き確率表

(0~100回:季節=夏, 曜日=×, 時間帯=朝, 天気=曇り)  
(101~200回:季節=×, 曜日=休日, 時間帯=昼, 天気=曇り)

季節	曜日	時間帯	天気	話題															
				野球	料理	釣り	学生生活	野球	料理	釣り	学生生活	野球	料理	釣り	学生生活	野球	料理	釣り	学生生活
春	×	朝	曇り	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
夏	×	朝	曇り	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
秋	×	朝	曇り	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
冬	×	朝	曇り	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
				0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25

表2の濃い青は季節、薄い青は天気、緑は時間帯、黒は話題のノードに関してそれぞれ区切られている。背景が白の値は基準値(0.25)であり、それよりも高いほど赤く、低いほど青で表現している。

##### 4.2.4 考察

0~100回までは曜日の情報を観測できなかったため、平日も休日も両方学習を行っている。また、季節=夏、時間帯=朝、天気=曇りという条件の下で、長男は野球と学生生活に興味を示すため、表2のような学習結果となった。101回~200回までは、季節の情報を観測できなかったため、春、夏、秋、冬全てで学習を行っている。また、時間帯=昼、天気=曇りという条件の下で、長男は野球と学生生活に興味を示すため、表2のような学習結果となった。野球と学生生活の確率が基準値と比べて高く、料理と釣りの確率が基準値と比べて低くなっている。これらの結果より、インタラクションの相手への話題提供に関する行動の学習ができていることが分かる。今回のシミュレーションでは、長男の興味をあらかじめ定義していたが、長男とのインタラクションを重ねる中で長男の興味に変化したとしても、それに応じた学習ができるため、他者の趣味、趣向の変化に対応が可能である。父親、母親、長女に対しても同様のシミュレーション(趣味、趣向は異なる)を行ったところ、同様に行動の学習が行えていることを確認できた。

## 5 おわりに

本研究では、自己組織化マップ(SOM)、ベイジアンネットワーク(BN)、Q学習、Soft-max行動選択法を用いて、分人モデルの構築を行った。計算機シミュレーションにより、自己組織化マップを用いてインタラクションの相手の特定と学習を行い、ベイジアンネットワーク、Q学習、Soft-max行動選択法を用いて、話題提供に関する行動決定と学習が行えることを確認した。この分人モデルによって、社会学における人のインタラクションに関する分野の性能に利用できることを期待している。

## 参考文献

- [1] 平野啓一郎:「私とは何か「個人」から「分人」へ」、講談社新書(2012)
- [2] T. コホネン(徳高平蔵他監修):「自己組織化マップ改訂版」、丸善出版(2012)
- [3] C.M. ビショップ(元田 浩 他監訳):「パターン認識と機械学習(下)」, 丸善出版(2012)
- [4] C. J. C. H. Watkins and P. Dayan: "Q-learning", Machine Learning, Vol.8, No.3-4, pp.279-292 (1992)
- [5] 大林 正直, 呉本 亮, 小林 邦和:「インテリジェントコンピュータディンク(改訂二版)」, 山口大学工学部知能情報工学科生体情報システム工学研究室(2012)